爬虫高级实战课程

2021年10月上线

2021 New Product Introduction

主讲人:樵夫

VISIT…

LANZAROTE Caliente.COM

01

课程主讲内容

课程在初学者的角度，从原理到实践，循序渐进地讲解Python开发网络爬虫的核心技术。全面介绍了数据采集，存储，动态网站爬取，验证码破解，模拟登陆，代理使用，爬虫框架，分布式爬取，JS逆向伪装等知识，精选多个实战项目，从易到难，层层深入。不同项目解决不同的抓取问题，带你从容抓取主流网站，让你牢牢掌握爬虫工程师硬核技能。

02 适用人群

1.零基础人群/热爱爬虫技术的小白

2.有一定python基础经验的人群

3.在校大学生人群（毕设）

03 课程亮点

名师精品教学

录播（3周）+直播（8周）教学模式，九大模块。对标全网性价比高，内容全套的课程大纲

爬虫高级技术解密

超详细的Scrapy框架案例式教学+超详细的JS逆向实战教学-让你不再担心所有JS数据加密

20+案例全程贯穿

根据不同网站反爬机制，3个月通过20+爬虫案例（图片/小说/视频/商业网站）手把手带学员逐一破解

案例永久更新

课程内案例爬取技术如果失效，案例代码持续更新

04 课程服务

01 学习周期

课程学习周期五年有效，后续更新爬虫项目免费分享

02 作业检测

9大学习模块配套考核作业，实时监督你的学习效果

03 答疑解惑

VIP群学员享受群答疑周期一年

04 跟踪督学

讲师+班主任+辅导老师全程跟踪。
实现保姆式教学体验模式！

05 保驾护航

持续分享学习、面试、工作中遇到的问题、困难和新的技术。

06 成长陪伴

后续推出的课程VIP学员可享受当时最低价的8折优惠。

品质服务

品质源于坚持

05 爬虫课程大纲

一模块、爬虫概念【录播】

1.1爬虫入门	本课程使用软件说明 第一个爬虫 Urllibweb 请求过程剖析剖析
1.2 Chrome浏览器功能	Elements Sources Console Network
1.3 Http协议	请求 响应 GET POST 状态码
1.4 requests模块入门	requests安装 requests发送get请求 多参数的get请求处理 requests发送post请求

二模块、数据解析（re，bs4，xpath，pyquery）【录播】

2.1正则表达式	元字符 量词 惰性匹配
2.2 re模块详解	findall() finditer() search() Match对象 group分组 compile()预加载
2.3 案例	豆瓣TOP250抓取案例
2.4 html基础语法规则	html基础语法规则
2.5 css基础语法规则	css引入方式 选择器 id选择器 类选择器 标签选择器 组合选择器 子选择器 后代选择器 属性选择器

二模块、数据解析（re，bs4，xpath，pyquery）【录播】

2.6 bs4模块安装与基本的使用	bs4模块安装 BeautifulSoup对象 find() find_all() 属性筛选 文本提取
2.7 案例	抓取优美图库的图片
2.8 xpath入门	lxml模块安装 节点概念 xpath处理XML xpath节点筛选 xpath属性筛选 xpath文本提取 xpath相对筛选 xpath处理HTML
2.9 案例	xpath案例-抓取猪八戒数据
2.10 pyquery入门	pyquery安装 pyquery基本语法 选择器的使用 pyquery链式提取 pyquery文本提取 pyquery操纵dom
2.11 案例	pyquery案例

三模块、初识反爬【录播】

3.1 cookie的使用	cookie是什么? cookie的作用是什么? 伪装cookie session的概念 利用session维护cookie 实战模拟登录伪装
3.2 referer的使用	防盗链的概念 referer的作用
3.3 案例	梨视频
3.4 代理的使用	代理的概念 代理如何使用 代理池 第三方代理接入

四模块、异步爬虫【直播】

4.1多线程	Thread类 start是怎么回事儿? 线程的状态变化
4.2 多进程	Process类 多线程与多进程对比
4.3 线程池	ThreadPoolExecutor类的使用 with在线程池的作用 submit()
4.4 案例	线程池实战案例-新发地菜价抓取

四模块、异步爬虫【直播】

4.5 协程	为什么要有协程? 协程应用场景 asyncio模块 async关键字 await关键字 多任务异步协程 aiohttp详解 ClientSession()详解 aiohttp数据提取 aiofile模块详解 利用协程抓取一部小说
4.6 电影抓取	视频网站的分析 M3U8文件下载 M3U8文件解析 ts文件批量下载 ts文件解密-AES解密 ts文件合并 爬虫自省-重点

五模块、自动化测试工具selenium【直播】

<table>
<tr><td>5.1 selenium概述与环境搭建</td><td>Chrome浏览器版本号确认
chromedriver版本确认
chromedriver下载与安装
selenium模块安装
selenium自动打开浏览器</td></tr>
<tr><td>5.2 selenium各种神奇操作</td><td>find_element()
by类型分析find_element_by_id()
find_element_by_tag_name()
find_element_by_class_name()
find_element_by_xpath()
find_elements()
click()
send_keys()
page_source
get_cookies()</td></tr>
</table>

5.3 多窗口调度与iframe处理	switch_to.window() window_handles switch_to.frame() select_by_index() select_by_value() select_by_visible_text() selenium处理滚动条 execute_script() 窗口大小调整
5.4 无头浏览器	等待时间设置 无头浏览器配置 Options() --headless --disable-gpu

五模块、自动化测试工具selenium【直播】

5.5 超级鹰使用	验证码问题 超级鹰账号注册 超级鹰官方示例解读 验证码识别类型 返回数据解读与提取 超级鹰来模拟登录超级鹰
5.6 模拟12306网站登录	12306页面分析 selenium截图_x0008_ 破解验证码 破解滑块 action动作链 防止selenium被浏览器检测到

六模块、持久化数据存储【直播】

6.1 MySQL数据库入门	数据库介绍 MySQL数据库安装 Navicat安装与破解 Table的相关操作 insert语句 update语句 delete语句 select语句 group by 与聚合函数 order by排序
6.2 pymysql快速入门	pymysql模块安装 pymysql链接数据库 Cursor pymysql增删改查操作 事务处理 抓取网吧电影所有电影信息存储到数据库

六模块、持久化数据存储【直播】

6.3 MongoDB数据库入门	MongoDB安装 添加数据 修改数据 删除数据 查询数据 复杂查询
6.4 pymongo模块	pymongo模块安装 pymongo链接MongoDB数据库 完成增删改查操作
6.5 案例	抓取二手房成交量存储到MongoDB与MySQL中

七模块、scrapy框架【直播】

7.1 Scrapy环境安装	scrapy简介 scrapy安装
7.2 Scrapy基本使用	scrapy全局命令 scrapy项目命令 创建项目 创建爬虫 编写爬虫 运行爬虫
7.3 Spider	name名称 allowed_domains start_urls start_requests重写起始方法 parse解析方式 Request请求对象详解

七模块、scrapy框架【直播】

7.4 Scrapy组件	spider编写爬虫逻辑 item定义数据格式 pipeline数据存储 下载中间件
7.5 Item数据模型	Item积累 Field字段类 扩展Item类
7.6 ItemPipeline	自定义管道 数据存储到csv 数据存储到MySQL 数据存储到MongoDB 图片，文件下载管道 同时处理多个Item

七模块、scrapy框架【直播】

7.7 下载中间件	process_request process_response 请求头反爬处理 动态UA IP反爬处理 cookie反爬处理
7.8 案例	Scrapy实战案例一
7.9 CrawlSpider全站数据爬取	LinkExtractors链接提取器 Rules爬取规则设置 Rule链接提取规则 CrawlSpider全站爬取

七模块、scrapy框架【直播】

7.10 过滤器	MD5算法 指纹过滤器 request_fingerprint requst_seen RFPDupeFilter
7.11 Scrapy配置文件	Robots协议配置 日志配置 重定向配置 下载延迟 管道，中间件优先级配置

7.12 分布式爬虫

- redis简介
- redis安装与配置
- redis数据类型
- redis数据增删改查
- pyredis快速上手
- 分布式概念
- 使用redis完成分布式爬虫
- RedisSpider
- 分布式持久化
- 分布式调度器
- 起始URL设置
- redis实现增量爬虫

七模块、scrapy框架【直播】

7.13 Scrapy-Splash	docker安装 splash拉取与安装 splash配置 Scrapy-Splash安装 利用Scarpy-Splash完成数据抓取
7.14 案例	scrapy实战案例二

八模块、Python高级反爬-js逆向【直播】

8.1JavaScript语法入门

变量

if判断

for循环

Switch

js函数

forEach()

闭包函数

JS对象

eval()

原型

Date与Math

定时器

js操纵DOM对象

Chrome调试JS代码

八模块、Python高级反爬-js逆向【直播】

8.2 各种加密算法	URLEncode编码解码 BASE64编码解码 MD5加密 SHA1 DES对称加密 AES加密 RSA非对称加密 某易音乐实战
8.3 PyExceJS模块	NodeJS安装 PyExecJS模块安装 复杂js自动化处理 PyExecJS调用js文件
8.4 JS逆向开发实战	实战-etw360登录解密 实战-完成战火登录 实战-完成空气质量数据批量抓取

九模块、JS逆向实战、APP逆向实战【直播】

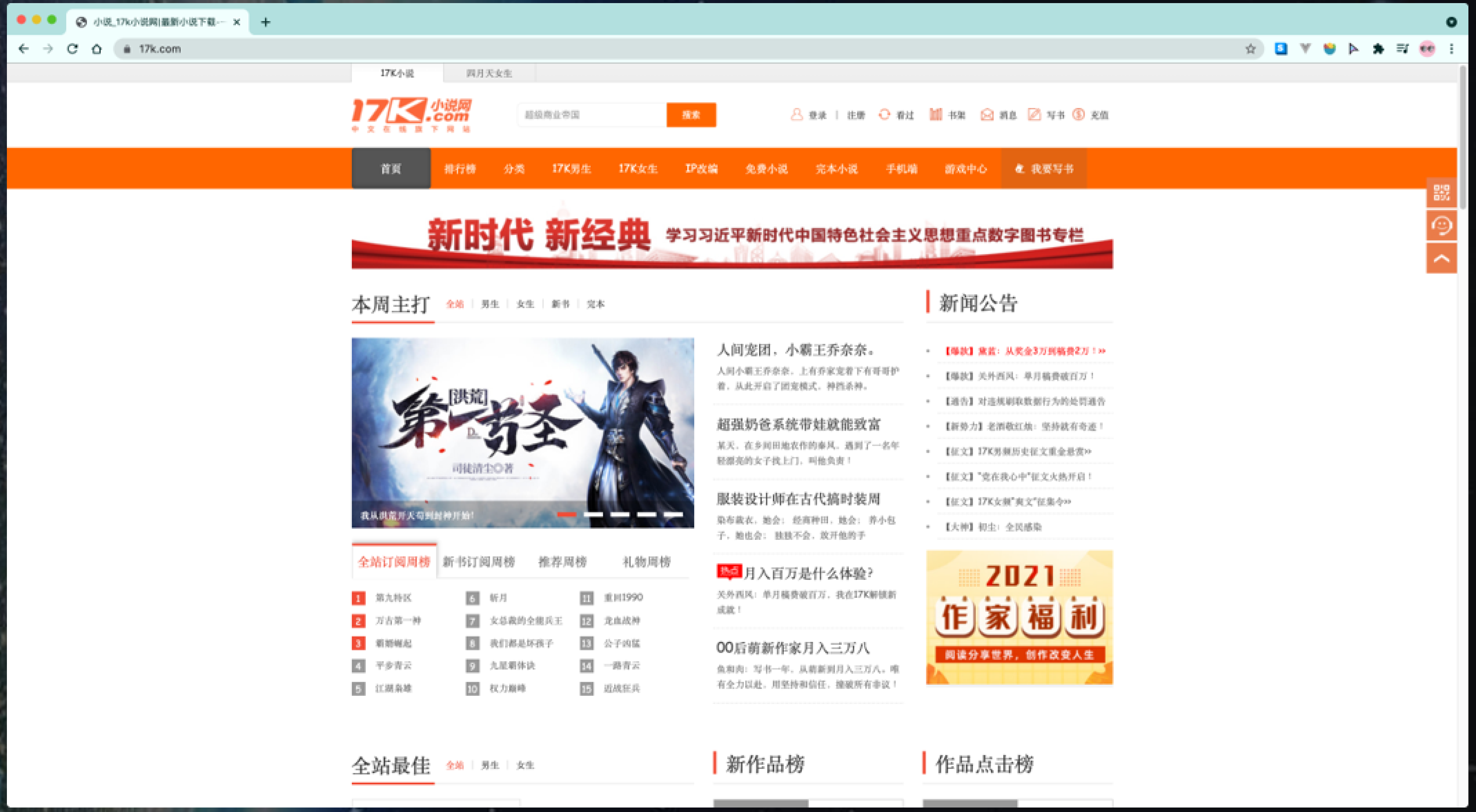
9.1 JS逆向实战	chrome断点调试 js代码结构分析 必备加密算法逆向和实现（案例核心算法） node.js编译原生js算法 asyncio、aiohttp异步编程 高阶进程&线程&协程提升并发量 python高性能异步框架uvloop 企业应用级的日志处理 某马逊 某视频播放量 某日头条 某音短视频
9.2 APP逆向实战	安卓模拟器 网络抓包和配置charles app脱壳 & 反编译

九模块、JS逆向实战、APP逆向实战【直播】

9.2 各种加密算法	基于Frida框架和xposed的Hook实战 Frida-RPC及Flask服务开发 Java必备语言基础 （逆向必备技能） 安卓APP和NDK开发（so破解） JNI静态分析和动态调试 AndroidNativeEmu和unidbg逆向so算法 某B站 某红书 某音短视频 某乎
9.3 基于Flask构建自动化爬虫平台	Flask框架 Flask原理剖析 MySQL和SQLAchemy 消息队列和缓存 基于Python实现exe爬虫应用脚本开发

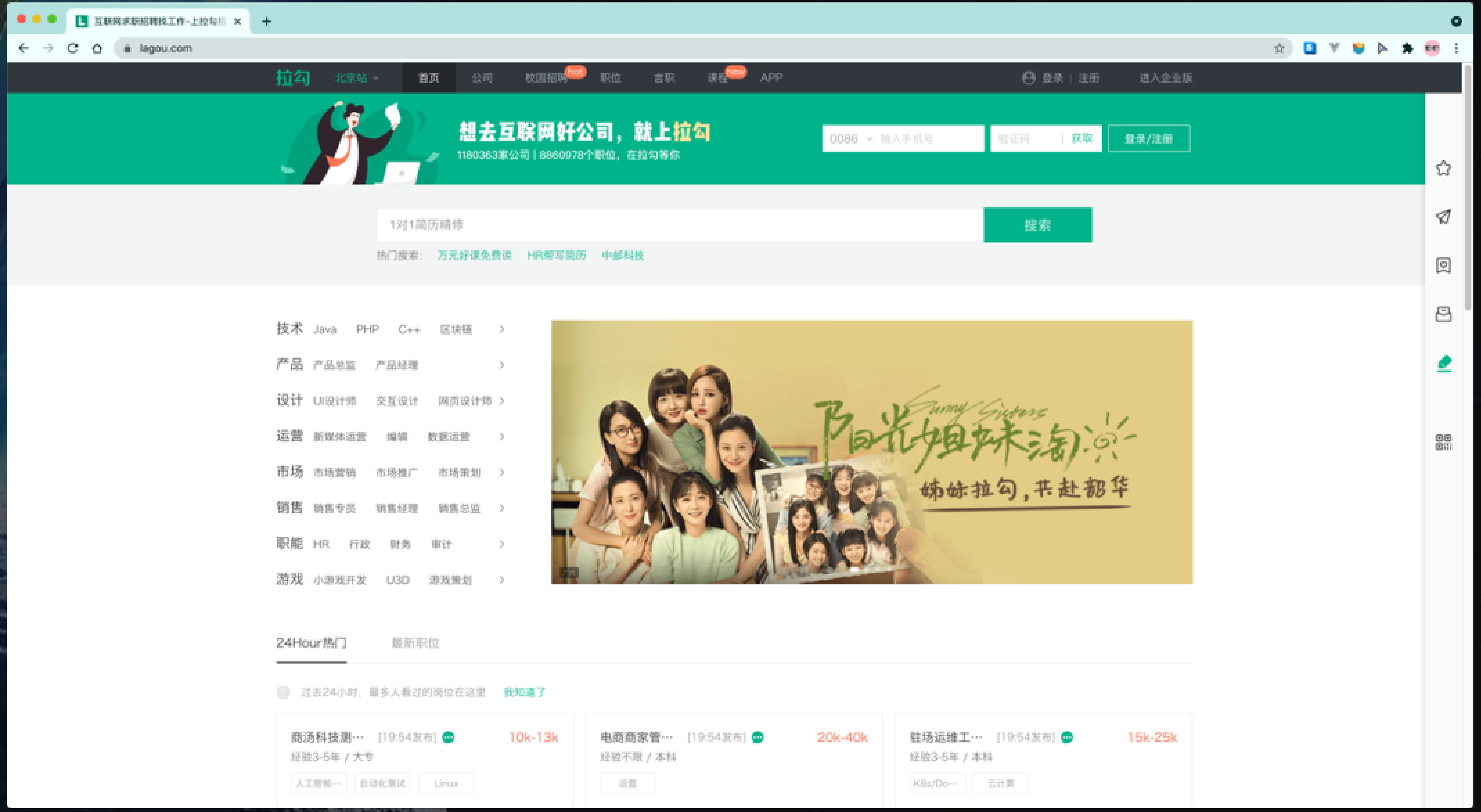
06

部分案例展示



17k小说抓取

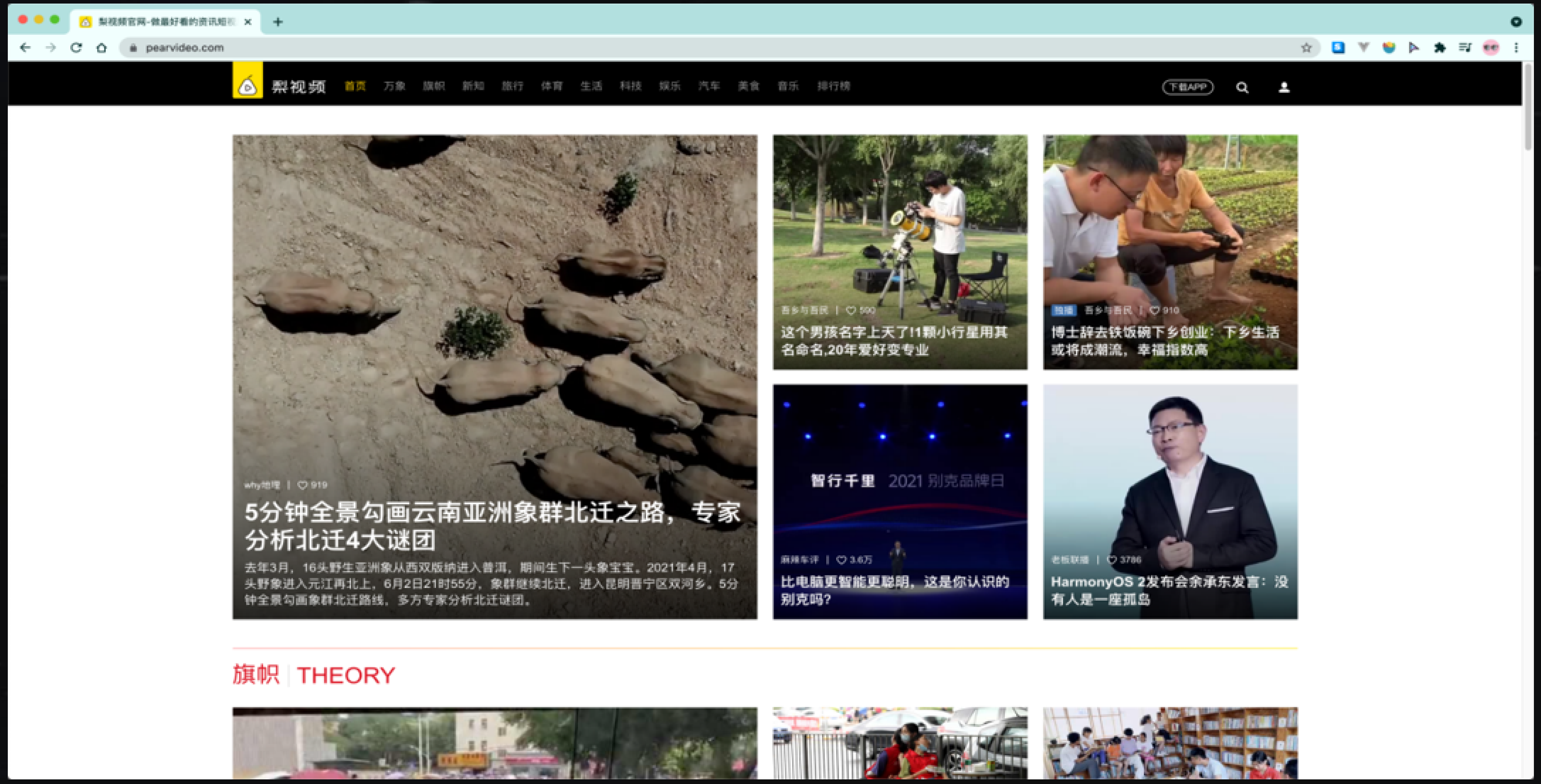
百度小说全集抓取



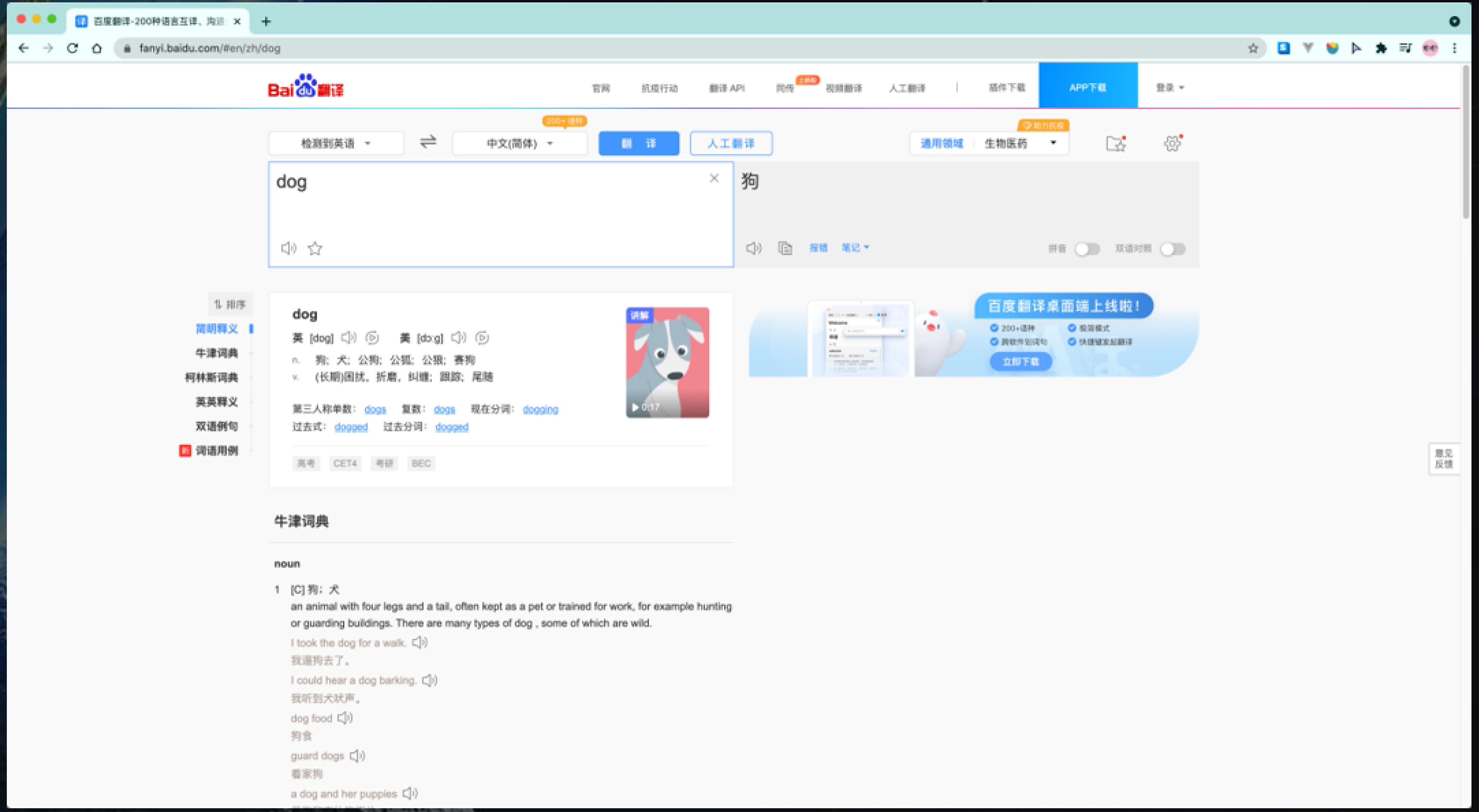
拉勾网招聘信息抓取

12306自动登录破解

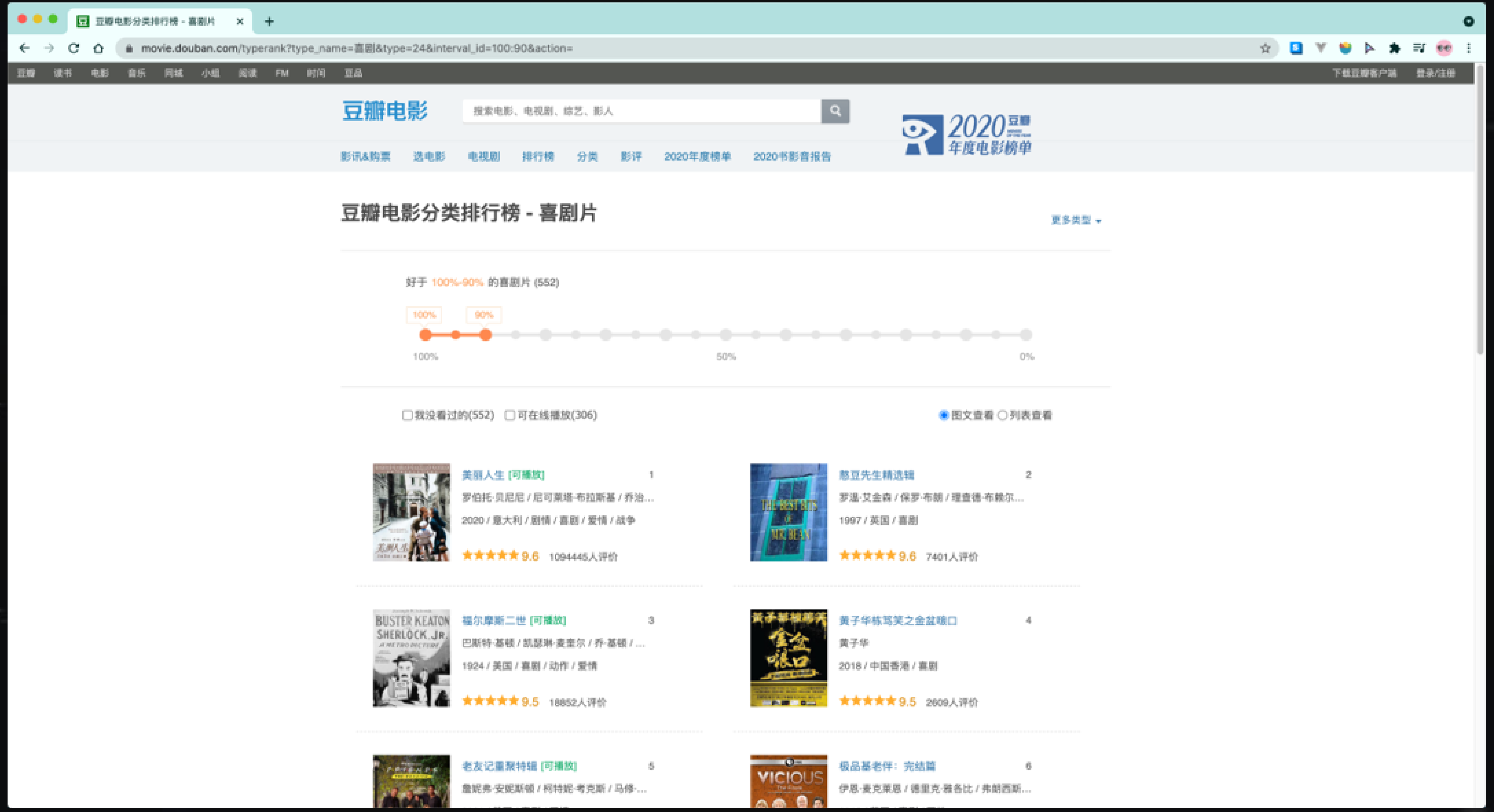
电影天堂电影信息抓取



梨视频-视频抓取

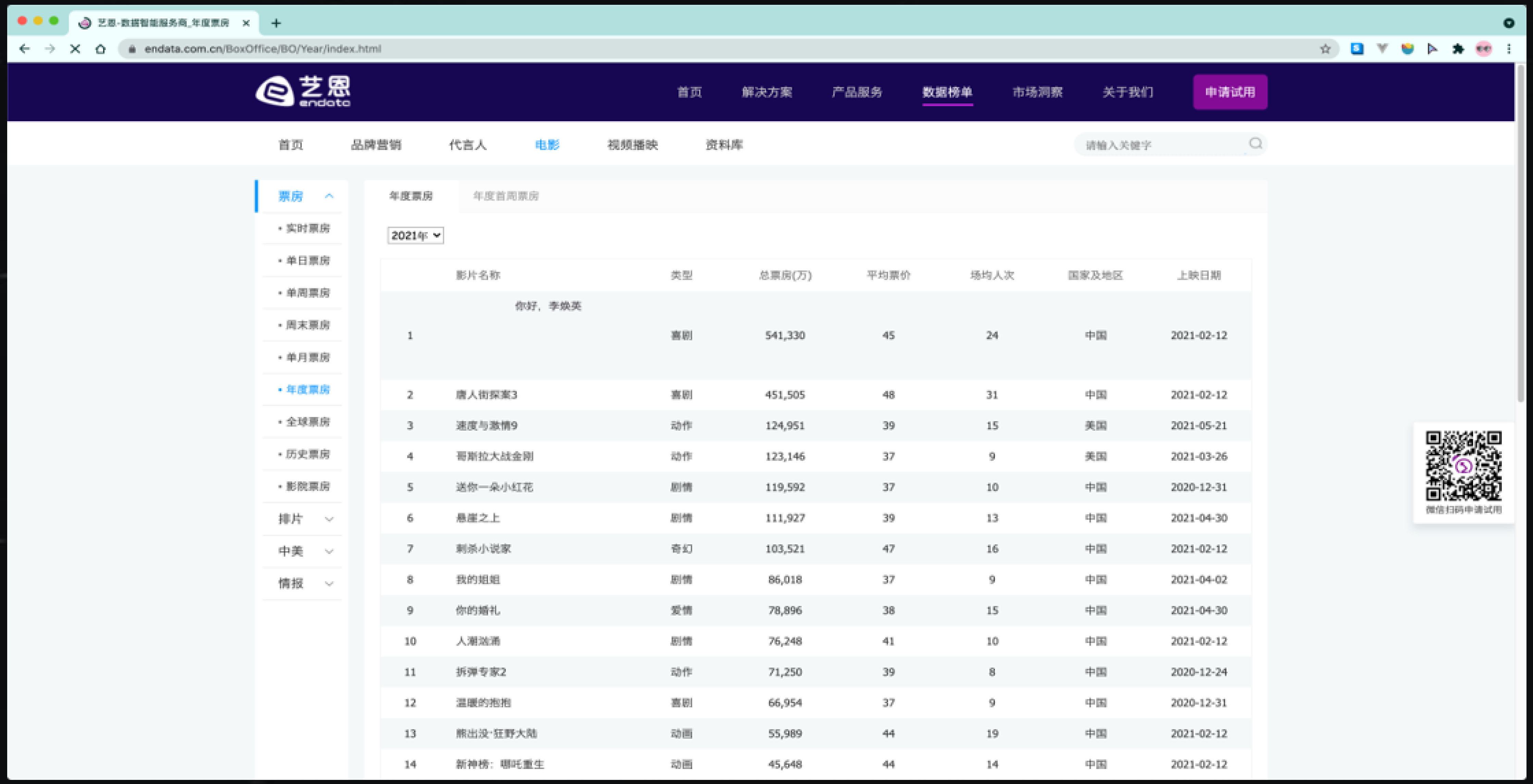


百度翻译部分数据提取

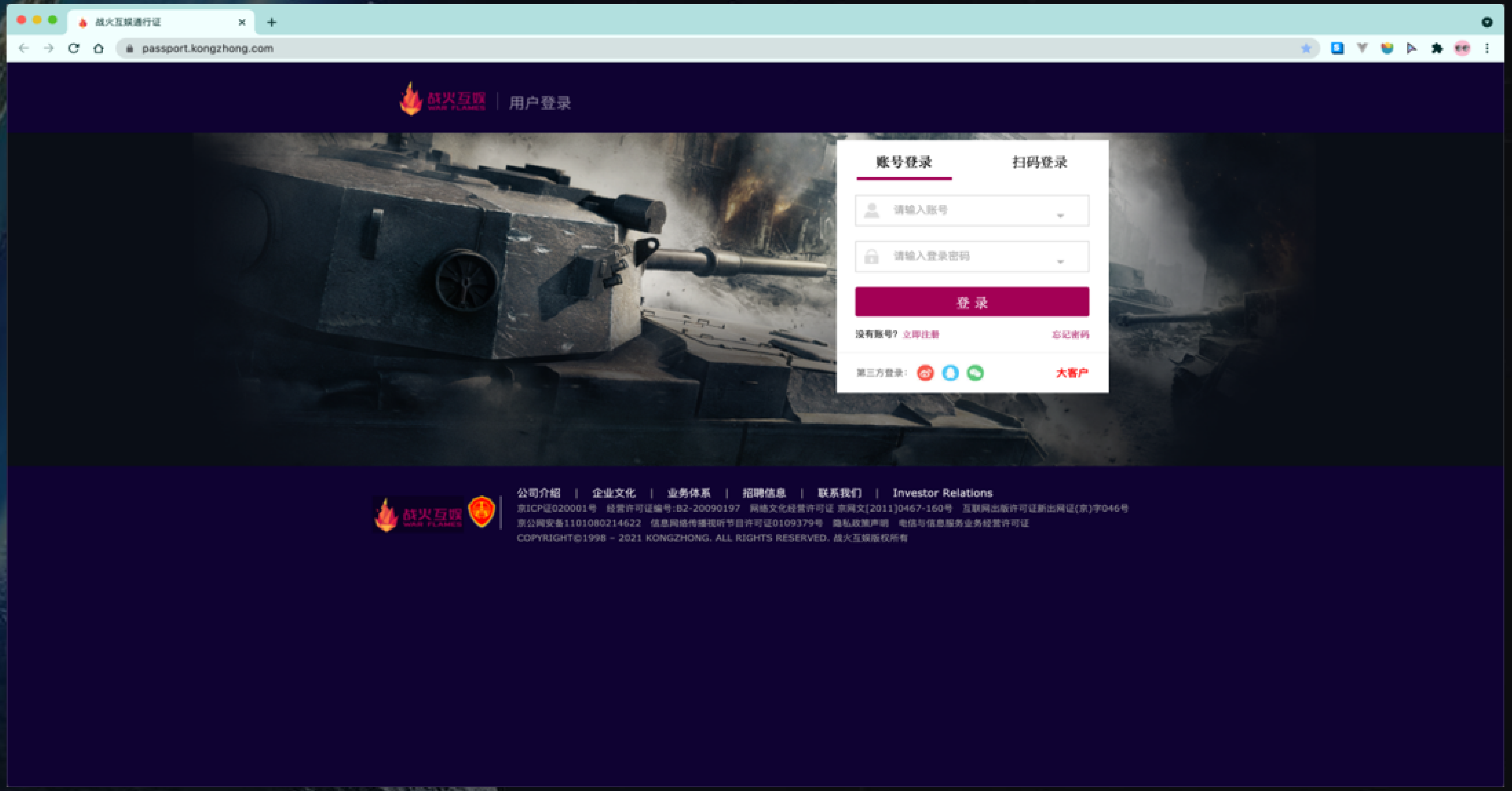


豆瓣电影排行榜-喜剧分类抓取

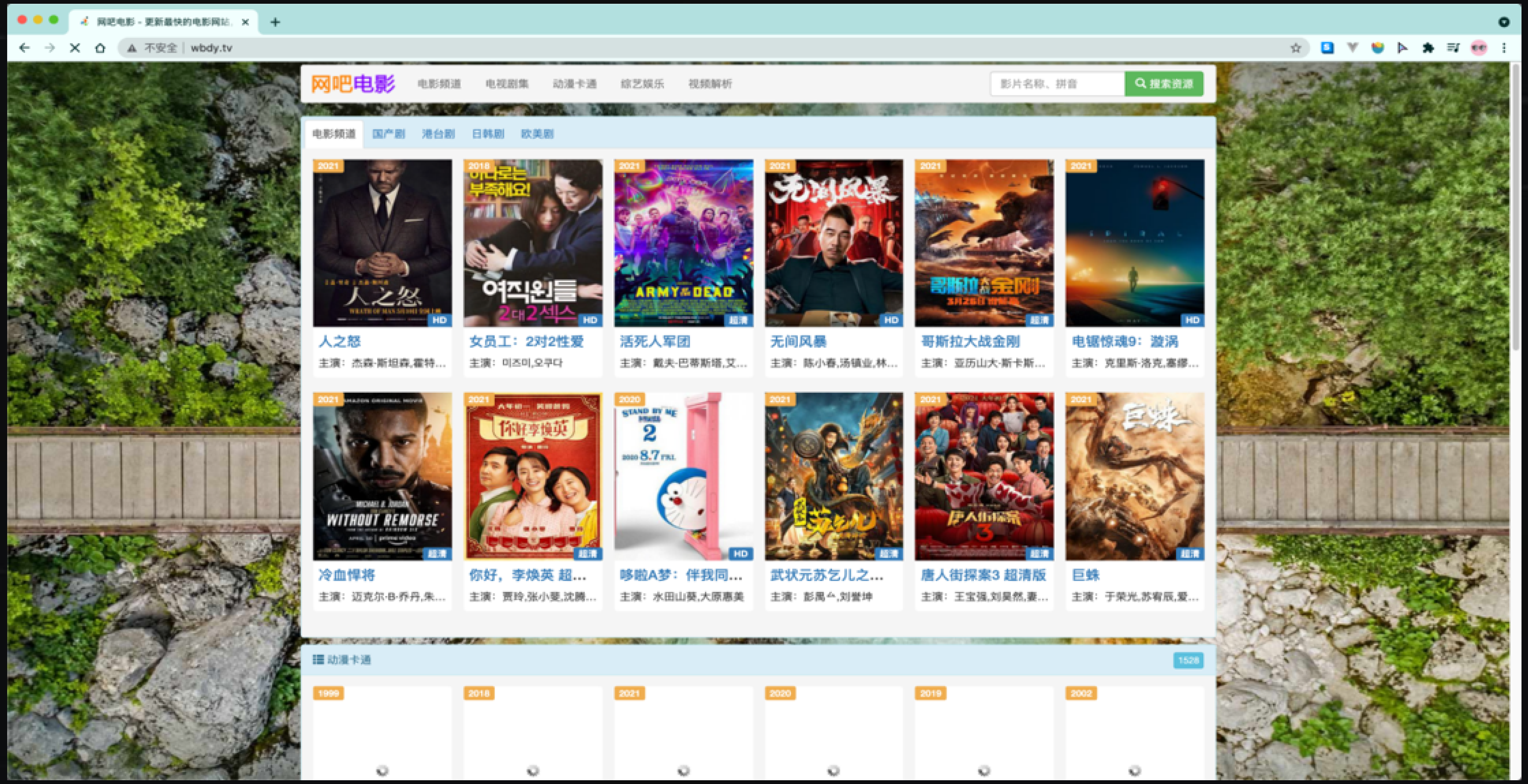
汽车之家-奥迪A8口碑信息



艺恩电影数据库抓取



战火互娱登录数据加密

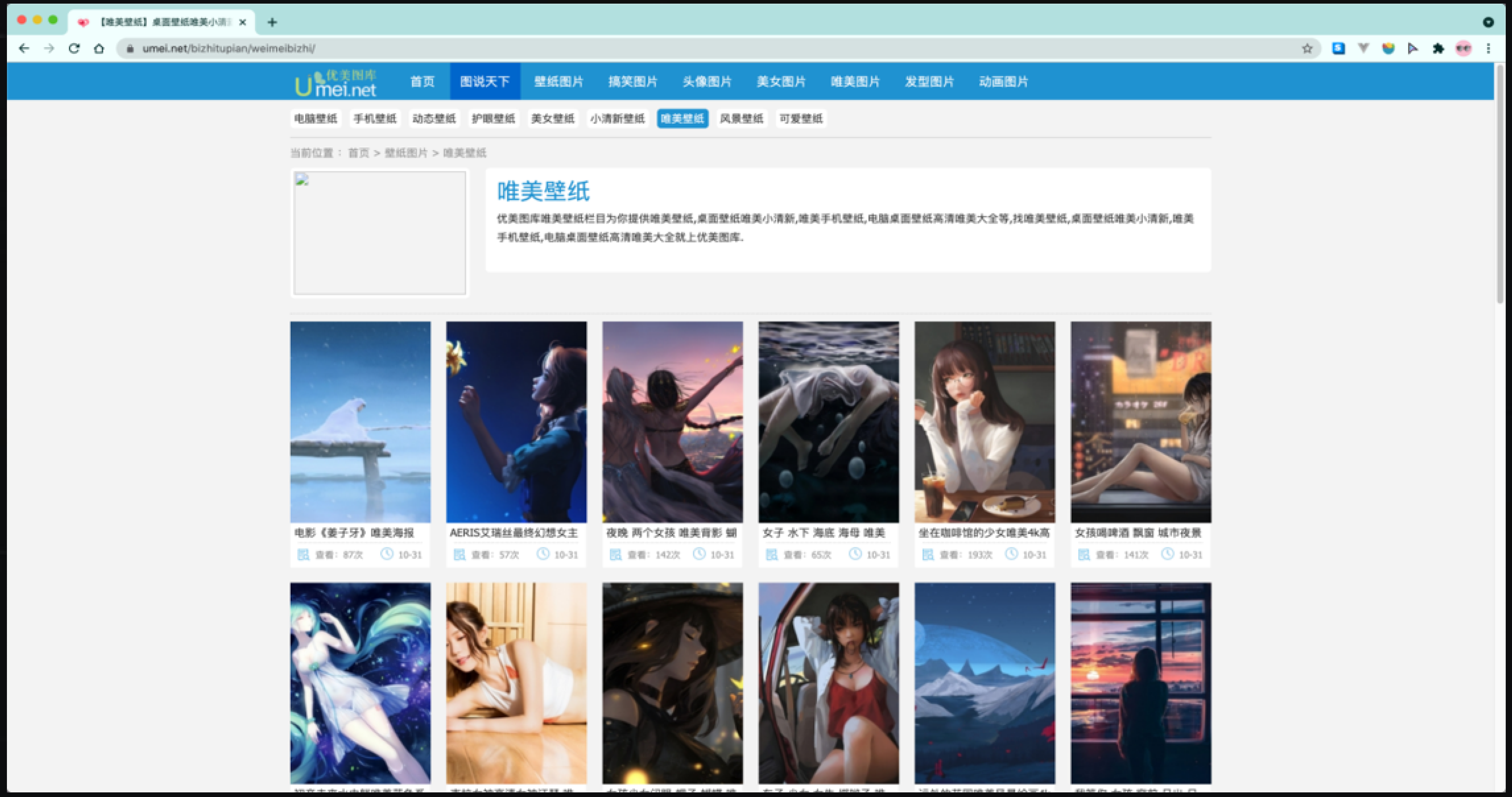


网吧电影-加密视频抓取

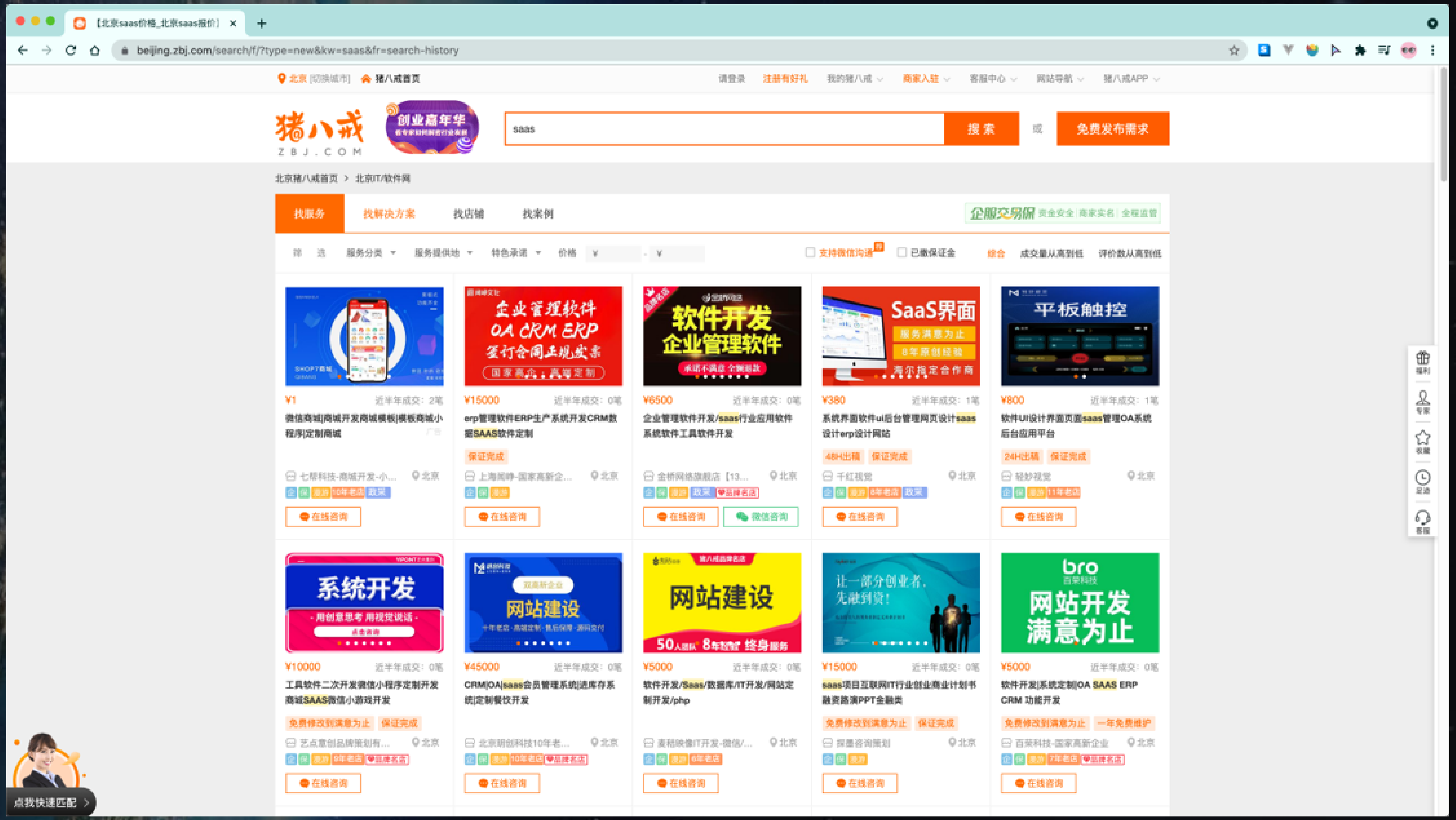
新发地菜价抓取

真气网加密数据破解

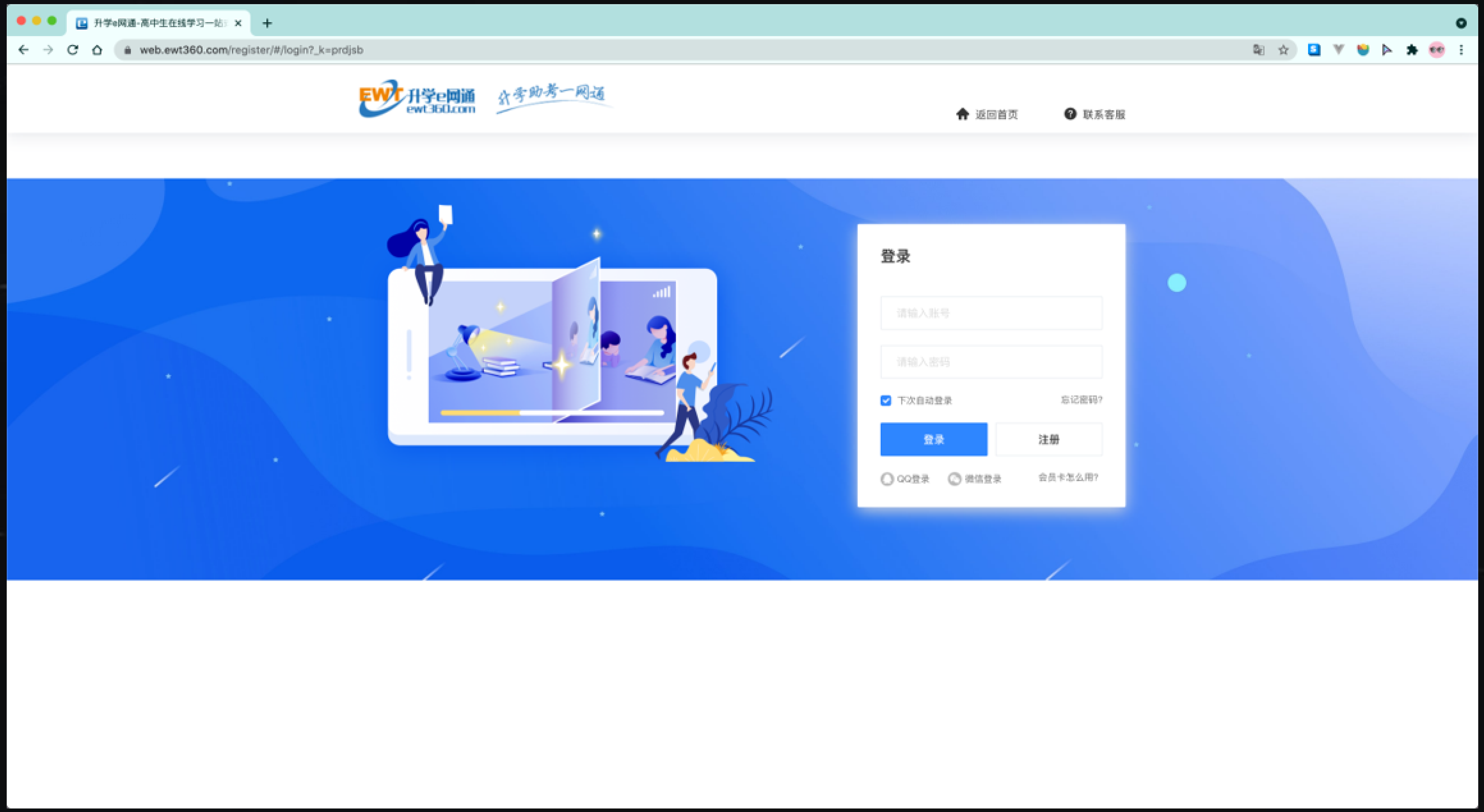
网易云音乐评论抓取



优美图库壁纸抓取



猪八戒服务信息抓取



etw360登录过程破解

中大网校考题抓取

真气网加密数据破解

07 往期好评

A****
★★★★★
怎么说好呢？只能这么讲了 卧槽，特么的 这是我目前见到的 讲的最好的爬虫课程 最主要的一点是， 老师真的很负责，经常晚上很晚了（晚上接近12点了），还在回复问题超级给力

已上课2天6时24分时评论　　2021-09-17

机构回复：怎么说好呢？只能这么讲了：感谢感谢

吴****
★★★★★
老师讲的太棒啦，很喜欢老师的讲课方式，虽然我一直是在看回播的路上，但出现问题也有老师的远程帮助，真的是美滋滋！！！加油！每一天

已上课1天15时42分时评论　　2021-08-22

机构回复：一起加油，每一天

范****
★★★★★
老师讲课幽默风趣，知识点很全面，开发思路清晰明了

已上课1天9时13分时评论　　2021-08-13

机构回复：感谢范同学，加油哦

腾讯课堂好评

明****
★★★★★
必须五星！老师太棒了！老师有没有数据分析的课程

已上课1天9时34分时评论　　2021-08-10

机构回复：数据分析预计要明年年初了，可以管你的班主任老师要一些今年新录的的数据分析的基础课程先看着哦。

T****
★★★★★
非常喜欢樵夫老师的课程！ 1.由浅至深，深浅有度。 深的课程剖析成易懂的语言来仔细讲，浅的课程拆开了掰碎了讲！ 重点课程分开讲，让每一位同学都能正常的通俗易懂的理解和明白其中的含义。 对于爬虫方面的知识深入骨髓。 2.有问必答，有疑必解。 在学习过程中遇到不懂得问题，樵夫老师会单独把这方面知识再次通过其他形式给你找齐。 3.以诚相待，全心办学 态度好！没有任何老师的架子，和同学们以诚相待。不像老师更像朋友！ 祝老师越办越好！！！

已上课16时2分时评论　　2021-07-29

机构回复：哈哈哈，感谢写了这么多。你最帅

腾讯课堂好评

五****
★★★★★
老师讲课十分幽默风趣，几乎没有翻车，讲的知识通俗易懂，连那些闭塞难懂的知识，都可以让我们简单易懂。

已上课15时20分时评论　　2021-07-30

机构回复：加油哦，马上就要到难点了，一定要跟住

追评：这课程不错，老师及时回复我们学习中的问题，还帮忙解决课后问题。

已上课2天54分时评论　　2021-09-11

****
★★★★★
B站上看的樵夫老师，讲的通俗易懂，有问题了会很负责任的讲解

已上课16时43分时评论　　2021-07-20

机构回复：哈哈哈 感谢感谢

1****
★★★★★
老师这儿课儿教得也忒棒嘞。。。

已上课21时23分时评论　　2021-09-01

腾讯课堂好评

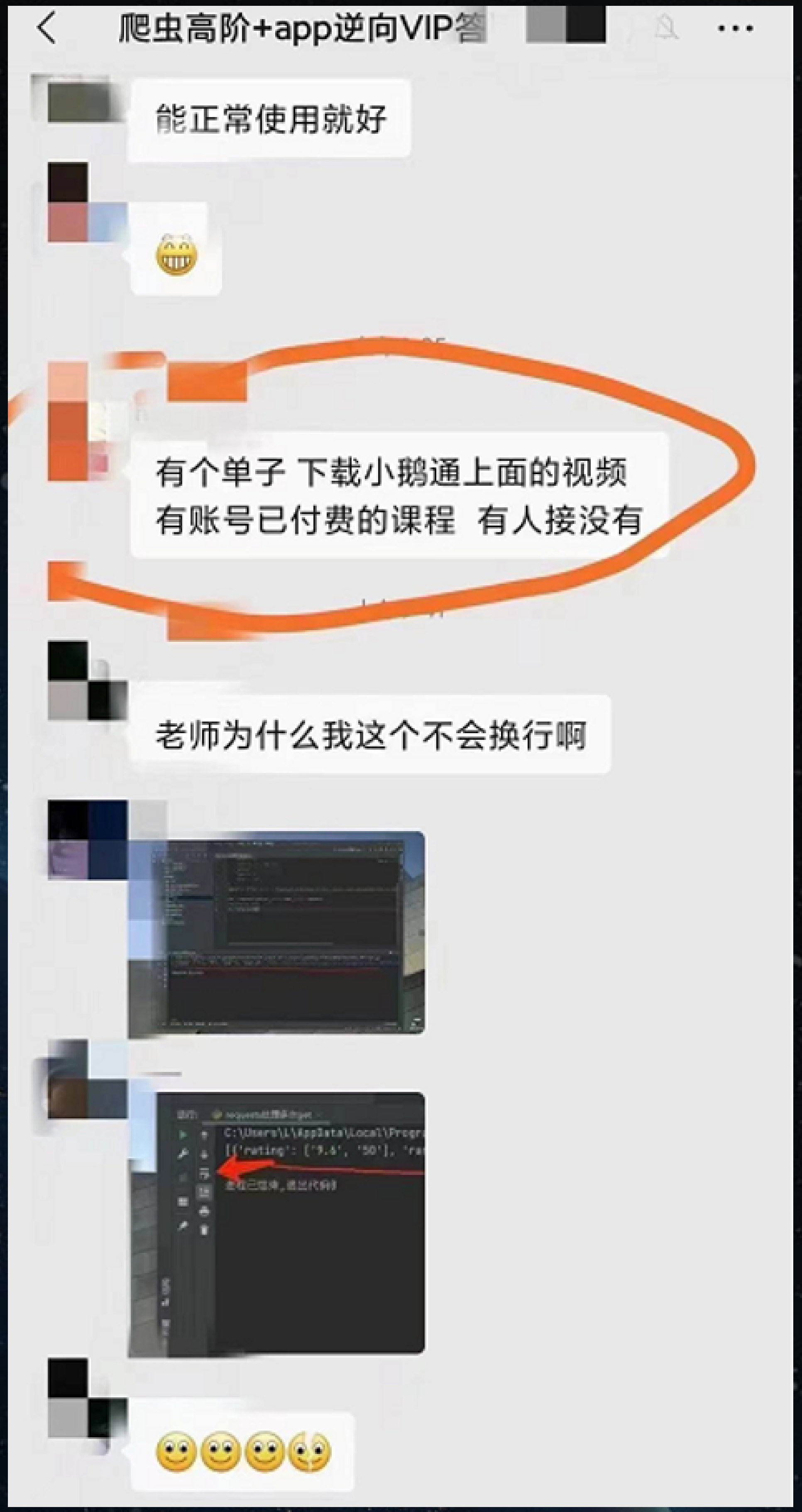
07 往期好评

腾讯课堂好评

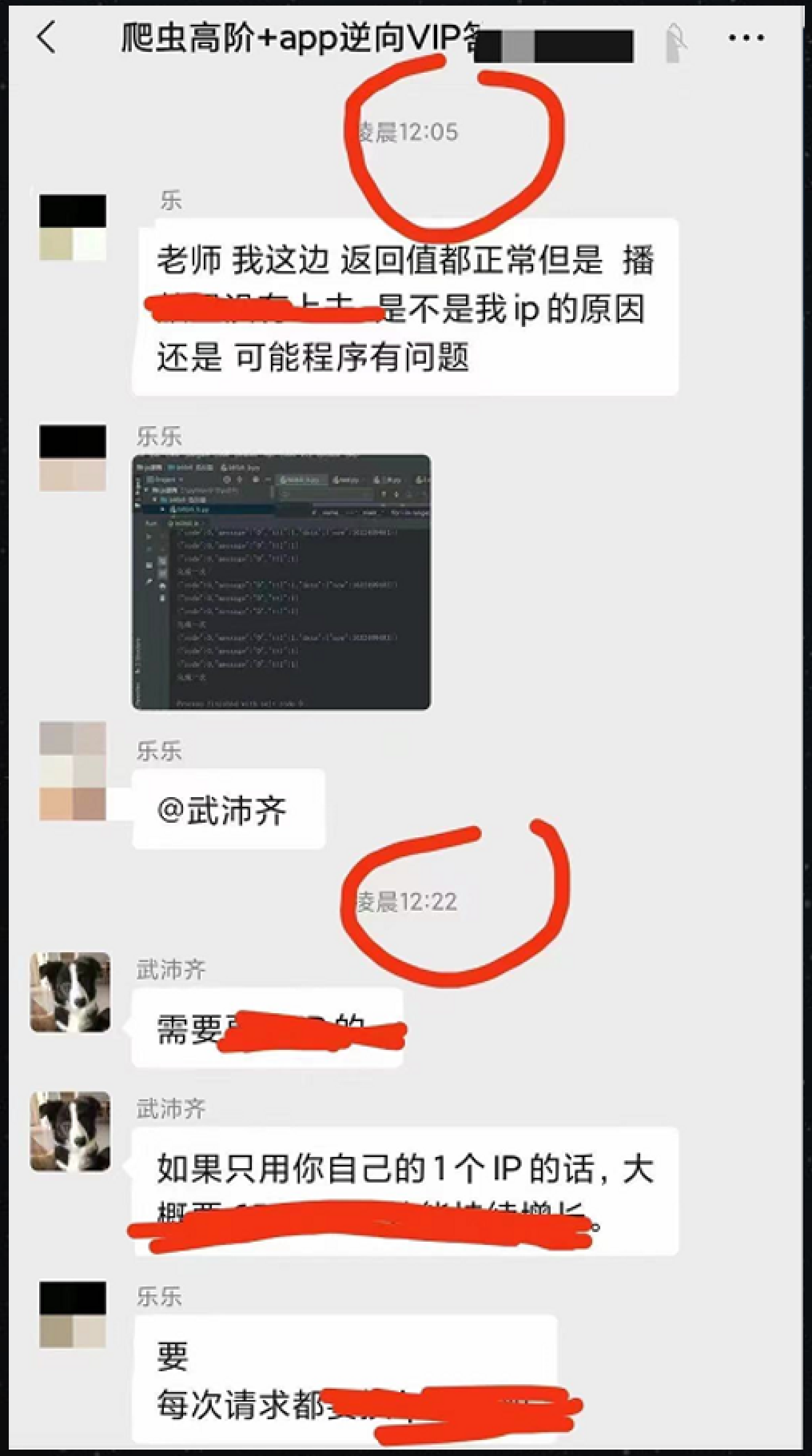
腾讯课堂好评

VIP群学员放单

07 往期好评



VIP群学员放单



VIP群夜晚答疑

VIP学员接单到手软

08 答疑解惑

Q：报名后悔了怎么办？
A：开课后报名7天内支持全额退费。
Q：报名后担心没有老师解答，报名前一个态度，报名后放任自身自灭？
A：我方被腾讯课堂邀请为联合运营机构，售后服务及老师技术门槛是最基本的联合保障。
付费课程配备小班级辅导群。群内班主任、辅导老师、主讲老师均在内答疑解惑。
Q：开课后还可以报名学习吗？
A：开课后依然可以报名学习。直播均在腾讯课堂自动生成回放视频。
Q：学费怎么交？
A：学费直接在腾讯课堂缴费即可。
Q：报名后如何和老师互动交流？
A：报名后有两种方式和老师互动：
1、在线直播实时互动交流，每次直播会预留答疑时间
2、班级学习群，群内有班主任、助教和讲师全方位答疑。
Q：课程一共多少课时？
A：课程一共录播35节，直播21节。每节直播至少2个小时。
一周平均直播2—3节（周二、四、六晚20:00-22:00授课），以腾讯课堂排课为准。
Q：课程如何授课？
A：课程采用录播+直播+答疑授课。通过知识讲解+项目实战+在线辅导的形式完成。